



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 128112**

**TO: Karen A Lacourciere**

**Location:**

**Art Unit: 1635**

**July 26, 2004**

**Case Serial Number: 09/301380**

**From: P. Sheppard**

**Location: Remsen Building**

**Phone: (571) 272-2529**

**sheppard@uspto.gov**

### **Search Notes**

Sheppard, Paula

18 2/12

**From:** Lacourciere, Karen  
**Sent:** Wednesday, July 21, 2004 3:48 PM  
**To:** Sheppard, Paula  
**Subject:** RE: Sequence search

Hi Paula-  
The numbers are  
ABA70229  
AAI50358  
AAK44359  
AAK18454  
ABS44016  
ABS18595  
ABA71469  
AAI51738  
AAK45811

The serial number is 09/301,380 and the sequence these oligos bind to is SEQ ID NO:1  
I can bring the original search down to you if it helps. Let me know what you need! Thanks so much, I really appreciate your help.  
Karen

-----Original Message-----

**From:** Sheppard, Paula  
**Sent:** Tuesday, July 20, 2004 5:28 PM  
**To:** Lacourciere, Karen  
**Subject:** RE: Sequence search

Yes. If you can provide the accession numbers of interest and the seq ID used to search it, then I can provide you with alignments and references.

Paula  
(571) 272-2529

-----Original Message-----

**From:** Lacourciere, Karen  
**Sent:** Tuesday, July 20, 2004 2:53 PM  
**To:** Sheppard, Paula  
**Subject:** Sequence search

Hi Paula-  
You did a sequence search for me back in May on SN 301,380. Some of the hits which appear in the summary now appear to be potentially useful references, but they are too far down on the list and the detailed summaries were not provided. Is it possible to get more information on these sequences based on the accession numbers?  
Thanks-  
Karen

*Karen A. Lacourciere Ph.D.*  
Remsen 2D15 GAU 1635  
(571) 272-0759

> O <  
O| | O Intelligence  
> O <

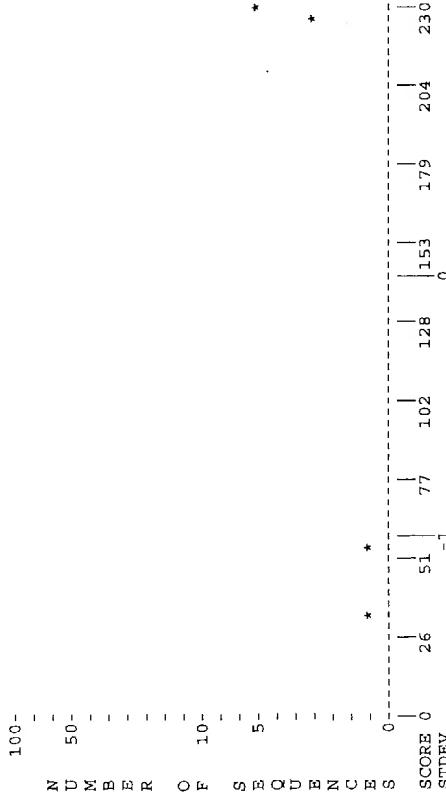
FastDB - Fast Pairwise Comparison of Sequences  
Release 5.4

Results file 1\_x\_listing\_inv.res made by spaula on Mon 26 Jul 104 14:24:17-PDT.

Query sequence being compared:US-09-301-380-1' (1-4134)  
Number of sequences searched: 8  
Number of scores above cutoff: 8

Results of the initial comparison of US-09-301-380-1' (1-4134) with:

File : aal50358.seq  
File : aal51738.seq  
File : aak18454.seq  
File : aak44359.seq  
File : aak45811.seq  
File : aba71469.seq  
File : abs18595.seq  
File : abs44016.seq



PARAMETERS

Similarity matrix Unitary K-tuple  
Mismatch penalty 1 Joining penalty 4  
Gap penalty 1.00 Window size 30  
Gap size penalty 0.33 500  
Cutoff score 0  
Randomization group 0

SEARCH STATISTICS

Scores: Mean 182 Median 225 Standard Deviation 84.92  
Times: CPU 00:00:00.90 Total Elapsed 00:00:00.00  
Number of residues: 2676  
Number of sequences searched: 8  
Number of scores above cutoff: 8

The scores below are sorted by initial score.

Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Score	Init. Opt.	Sig. Frame
1. aak18454	Human brain expressed single	230	230	230	0.57
2. aak44359	Human bone marrow expressed s	230	230	230	0.57
3. abs18595	Human genome-derived single e	230	230	230	0.57
4. abs44016	Human liver single exon probe	230	230	230	0.57
5. aak45811	Human bone marrow expressed s	224	224	224	0.49
6. aba71469	Human foetal liver single exo	224	224	224	0.49
7. aal50358	TOIG of: aal50358 check: 446	611	55	289	-1.50
8. aal51738	TOIG of: aal51738 check: 878	697	35	287	-1.73

1. US-09-301-380-1' (1-4134)  
aak18454 Human brain expressed single exon probe SEQ ID NO:

TOIG of: aak18454 check: 7397 from: 1 to: 230

ID AAK18454 standard; DNA; 230 BP.  
XX AC AAK18454;  
XX DT 05-NOV-2001 (first entry)  
XX DE Human brain expressed single exon probe SEQ ID NO: 18445.  
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;  
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
XX OS Homo sapiens.  
XX PN WO200157275-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000667.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX PI MPI; 2001-483446/52.  
XX DR Single exon nucleic acid probes for analyzing gene expression in human  
XX PT brains.  
XX PS Example 4; SEQ ID NO 18445; 650pp + Sequence Listing; English.  
XX CC The present invention provides a number of single exon nucleic acid  
XX CC probes which are derived from genomic sequences expressed in the human  
XX CC brain. They can be used to measure gene expression in brain cell samples,  
XX CC which may enable the diagnosis and improved treatment of nervous system  
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
XX CC epilepsy and cancers. The present sequence is one of the probes of the  
XX CC invention

XX Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

AAK18454 Length: 230 July 26, 2004 14:10 Type: N Check: 7397 ..

Initial Score = 230 Optimized Score = 230 Significance = 0.57  
Residue Identity = 100% Matches = 230 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

AAACATCTAGAGAAATGGAATATTGGCTTTGAGCTTAAATAAATAAAGGAATTCATGCGGTGACAGG  
70 80 90 100 110 120 130  
AGGAATTCATGCGGTGACAGG  
X 10 20

140 150 160 170 180 190 200  
AGAGGTGCGCTCTGAGCTTCCTTCCTTCAGCGGCTCTCTCTTTTACCACTGATGTCCTCAATAAA  
AGAGGTGCGCTCTGAGCTTCCTTCCTTCAGCGGCTCTCTCTTTTACCACTGATGTCCTCAATAAA  
30 40 50 60 70 80 90

210 220 230 240 250 260 270  
GGAGCCATCTCTCATTTGAATGAGGCTTAAACCCCTTCTCCATAGTCAACTAGGCTGCTCTCTCTTT  
GGAGCCATCTCTCATTTGAATGAGGCTTAAACCCCTTCTCCATAGTCAACTAGGCTGCTCTCTCTTT  
100 110 120 130 140 150 160

280 290 300 310 320 330 340 350  
TTTCACAGTCTCTGCTGAGGAGTTTCGACTTCCTTTTCAAGGCTTTGCTCTCTGCTCACTGTATTC  
TTTCACAGTCTCTGCTGAGGAGTTTCGACTTCCTTTTCAAGGCTTTGCTCTCTGCTCACTGTATTC  
170 180 190 200 210 220 230

TCCAAATGTCCTCATCTCTCTTCATAGGCTGGATTTCAGG

2. US-09-301-380-1' (1-4134)  
aak44359 Human bone marrow expressed single exon probe SEQ

TOIG of: aak44359 check: 7397 from: 1 to: 230

ID AAK44359 standard; DNA; 230 BP.

AC AAK44359;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 18916.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WC200157276-A2.

XX PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US0000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.

XX Example 4; SEQ ID NO 18916; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC bone marrow. They can be used to measure gene expression in bone marrow  
CC samples, which may enable the improved diagnosis and treatment of cancers  
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of  
CC the probes of the invention

XX Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

AAK44359 Length: 230 July 26, 2004 14:10 Type: N Check: 7397 ..

Initial Score = 230 Optimized Score = 230 Significance = 0.57  
Residue Identity = 100% Matches = 230 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

AAACATCTAGAGAAATGGAATATTGGCTTTGAGCTTAAATAAATAAAGGAATTCATGCGGTGACAGG  
70 80 90 100 110 120 130  
AGGAATTCATGCGGTGACAGG  
X 10 20

140 150 160 170 180 190 200  
AGAGGTGCGCTCTGAGCTTCCTTCCTTCAGCGGCTCTCTTTTACCACTGATGTCCTCAATAAA  
AGAGGTGCGCTCTGAGCTTCCTTCCTTCAGCGGCTCTCTTTTACCACTGATGTCCTCAATAAA  
30 40 50 60 70 80 90

210 220 230 240 250 260 270  
GGAGCCATCTCTCATTTGAATGAGGCTTAAACCCCTTCTCCATAGTCAACTAGGCTGCTCTCTCTTT  
GGAGCCATCTCTCATTTGAATGAGGCTTAAACCCCTTCTCCATAGTCAACTAGGCTGCTCTCTCTTT  
100 110 120 130 140 150 160

280 290 300 310 320 330 340 350  
TTTCACAGTCTCTGCTGAGGAGTTTCGACTTCCTTTTCAAGGCTTTGCTCTCTGCTCACTGTATTC  
TTTCACAGTCTCTGCTGAGGAGTTTCGACTTCCTTTTCAAGGCTTTGCTCTCTGCTCACTGTATTC  
170 180 190 200 210 220 230

TCCAAATGTCCTCATCTCTCTTCATAGGCTGGATTTCAGG

3. US-09-301-380-1' (1-4134)  
abs18595 Human genome-derived single exon probe ORF from lu

TOIG of: abs18595 check: 7397 from: 1 to: 230

ID ABS18595 standard; DNA; 230 BP.

AC ABS18595;

XX 19-AUG-2002 (first entry)

XX Human genome-derived single exon probe ORF from lung SEQ ID No 18586.

XX Human; ds; single exon probe; asthma; lung cancer; COPD; IID;

XX chronic obstructive pulmonary disease; interstitial lung disease;

XX familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX primary ciliary dyskinesia; pulmonary hypertension;

XX hyaline membrane disease; open reading frame; ORF.

XX Homo sapiens.

XX PN W0200186003-A2.  
XX PD 15-NOV-2001.  
XX PF 30-JAN-2001; 2001WO-US000665.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2002-114183/15.  
XX PT Spatially-addressable set of single exon nucleic acid probes, used to  
XX measure gene expression in human lung samples.  
XX PS Claim 4; SEQ ID NO 18586; 63app; English.  
XX CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a single exon probe open reading frame of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

ABS18595 Length: 230 July 26, 2004 14:11 Type: N Check: 7397 ..  
Initial Score = 230 Optimized Score = 230 Significance = 0.57  
Residue Identity = 100% Matches = 230 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

70 80 90 100 110 120 130  
AAACATTTAGAGAAATGGAATATTGGCTTTGAGCTTAAATAATTAACAAGGAATTCATGGCGTTGACAGG  
|||||  
AGGAAATTCATGGCGTTGACAGG  
X 10 20  
140 150 160 170 180 190 200  
AGAGGTGCTCTGAGCTTTTCCTTTCAGCCGGCTCTTTCTCTTTTACCACCTGATTTGTCACATAAA  
|||||  
AGAGGTGCTCTGAGCTTTTCCTTTCAGCCGGCTCTTTCTCTTTTACCACCTGATTTGTCACATAAA  
30 40 50 60 70 80 90  
210 220 230 240 250 260 270  
GGAGGCATCCTCATTTGAACCTGGCCATTAAACCCCTTCTCATAGTCAACATAGGCTGTCGTCACATCTTT  
|||||  
GGAGGCATCCTCATTTGAACCTGGCCATTAAACCCCTTCTCATAGTCAACATAGGCTGTCGTCACATCTTT  
100 110 120 130 140 150 160  
280 290 300 310 320 330 340 X 350  
TTTTCAGTCCCTGCTGAGGAGTTGAGCTTCTTTTTCAAAGGCTTGTGCTTCTGTCACATCTTT  
|||||  
TTTTCAGTCCCTGCTGAGGAGTTGAGCTTCTTTTTCAAAGGCTTGTGCTTCTGTCATCA  
170 180 190 200 210 220 230  
TCCAAATGCCCATCATCTTCTTCATAGGCTGGATTTCAGG  
360 370 380 390  
4. US-09-301-380-1' (1-4134)  
abs44016 Human liver single exon probe, SEQ ID NO 19006.  
TOIG of: abs44016 check: 7397 from: 1 to: 230  
ID ABS44016 standard; DNA; 230 BP.  
XX AC ABS44016;  
XX AC ABS44016;  
XX DT 25-FEB-2003 (first entry)  
XX DE Human liver single exon probe, SEQ ID NO 19006.  
XX KW Human; single exon nucleic acid probe; liver; cirrhosis;  
KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
KW coronary heart disease; ss.  
XX OS Homo sapiens.  
XX PN W0200157273-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000664.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX DR WPI; 2001-488898/53.  
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX gene expression in human adult liver.  
XX PS Claim 4; SEQ ID NO 19006; 658pp; English.  
XX CC The invention relates to a single exon nucleic acid probe (SENp) (I) for

CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (1) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABS25011-ABS51005 represent human  
CC liver single exon nucleic acid probes of the invention. Note: The  
CC sequence information for this patent does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

SQ Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

ABS44016 Length: 230 July 26, 2004 14:11 Type: N Check: 7397

Initial Score = 230 Optimized Score = 230 Significance = 0.57  
Residue Identity = 100% Matches = 230 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

70 80 90 100 110 120 130  
AAACATCTAGAGAAATGGAATATGGCTTTGAGCTTAAATAATTAACAAGGAATTCATGGCGTTGACAGG  
|||||  
AGGAATTCATGGCGTTGACAGG  
X 10 20

140 150 160 170 180 190 200  
AGAGGTGCTCTGAGCTTTCCTTCCTGACCGGCTCTTCTTCTTACACATGATTTGCCAATAAA  
|||||  
AGAAAGTCTCTGAGCTTTCCTTCCTTCCTGACCGGCTCTTCTTCTTACACATGATTTGCCAATAAA  
30 40 50 60 70 80 90

210 220 230 240 250 260 270  
GGAGCCATCTCATGTAAGTGGCATTAAACCCCTTCCATAGTAAGTGGCTGCTACTATCTCTTT  
GGAGCCATCTCATGTAAGTGGCATTAAACCCCTTCCATAGTAAGTGGCTGCTACTATCTCTTT  
100 110 120 130 140 150 160

280 290 300 310 320 330 340 350  
TTTCACAGTCTGTCGAGGAGTTCGACTTCCTTTTCAAGGCTTGCTGCTCTGCTACTGTATTC  
|||||  
TTTCACAGTCTGTCGAGGAGTTCGACTTCCTTTTCAAGGCTTGCTGCTCTGCTACTGTATTC  
170 180 190 200 210 220 230

TCCAAATGTCCTCATCTTCTTCATAGGCTGGATTTTCAGG

5. US-09-301-380-1' (1-4134)

aak45811 Human bone marrow expressed single exon probe SEQ

TOIG of: aak45811 check: 7602 from: 1 to: 224

ID AAK45811 standard; DNA; 224 BP.

XX AAK45811;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 20368.

XX Human; bone marrow expressed exon; gene expression analysis; probe;  
XX microarray; cancer; leukaemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
PA Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488900/53.  
XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human bone marrow.  
PT

XX Example 4; SEQ ID NO 20368; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
XX probes which are derived from genomic sequences expressed in the human  
XX bone marrow. They can be used to measure gene expression in bone marrow  
XX samples, which may enable the improved diagnosis and treatment of cancers  
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of  
XX the probes of the invention

SQ Sequence 224 BP; 58 A; 59 C; 49 G; 58 T; 0 U; 0 Other;

AAK45811 Length: 224 July 26, 2004 14:11 Type: N Check: 7602

Initial Score = 224 Optimized Score = 224 Significance = 0.49  
Residue Identity = 100% Matches = 224 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0

1450 1460 1470 1480 1490 1500 1510  
TACTGTTCACCAATTCACAGCAGTTCCAGAGCCACCATTTGGAGGTTCTTCCAGATGTCCCATGA  
|||||  
TCTTCTCCAGATGTCCCATGA  
X 10 20

1520 1530 1540 1550 1560 1570 1580  
CTACAGCTGGCTCGGGGCAAAACCCATGTCATTCAGGGCTGAACCTTGTATGATGATGAACAAAGGTTG  
|||||  
CTACAGCTGGCTCGGGGCAAAACCCATGTCATTCAGGGCTGAACCTTGTATGATGATGAACAAAGGTTG  
30 40 50 60 70 80 90

1590 1600 1610 1620 1630 1640 1650  
GCGTGCTGAGACAAATATTTGGATACATTTGCCAACACACAGATGTCATTCATCATCACCATCTTTCT  
|||||  
GCGTGCTGAGACAAATATTTGGATACATTTGCCAACACACAGATGTCATTCATCATCACCATCTTTCT  
100 110 120 130 140 150 160

1660 1670 1680 1690 1700 1710 1720 1730  
GGGCCAGCTTAACCTTTGATGAGAGGCTGCCCCATTAAGATTCGAAACCATTCAGGCTTCCAGTAATCT  
|||||  
GGGCCAGCTTAACCTTTGATGAGAGGCTGCCCCATTAAGATTCGAAACCATTCAGGCTTCCAGTA  
170 180 190 200 210 220 X

CCAAATATCAGGCTCTGATCCAGTCTTCCACAG

6. US-09-301-380-1' (1-4134)

aba71469 Human foetal liver single exon nucleic acid probe

TOIG of: aba71469 check: 7602 from: 1 to: 224

ID ABA71469 standard; DNA; 224 BP.

XX ABA71469;

XX 01-FEB-2002 (first entry)

XX DE Human foetal liver single exon nucleic acid probe #19774.  
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX OS Homo sapiens.  
XX FN WO200157277-A2.  
XX PD 09-AUG-2001.  
XX PF 30-JAN-2001; 2001WO-US000669.  
XX PR 04-FEB-2000; 2000US-0180312P.  
XX PR 26-MAY-2000; 2000US-0207456P.  
XX PR 30-JUN-2000; 2000US-00608408.  
XX PR 03-AUG-2000; 2000US-00632366.  
XX PR 21-SEP-2000; 2000US-0234687P.  
XX PR 27-SEP-2000; 2000US-0236359P.  
XX PR 04-OCT-2000; 2000GB-00024263.  
XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-483447/52.  
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing  
XX PT gene expression in human fetal liver.  
XX PS Claim 4; SEQ ID NO 19774; 639pp + Sequence Listing; English.  
XX CC The invention relates to a single exon nucleic acid probe for measuring  
XX CC human gene expression in a sample derived from human foetal liver. The  
XX CC single exon nucleic acid probes may be used for predicting, measuring and  
XX CC displaying gene expression in samples derived from human fetal liver. The  
XX CC present sequence is a single exon nucleic acid probe of the invention.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 224 BP; 58 A; 59 C; 49 G; 58 T; 0 U; 0 Other;

ABA71469 Length: 224 July 26, 2004 14:11 Type: N Check: 7602 ..  
Initial Score = 224 Optimized Score = 224 Significance = 0.49  
Residue Identity = 100% Matches = 224 Mismatches = 0  
Gaps = 0 Conservative Substitutions = 0  
1450 1460 1470 1480 1490 X 1500 1510  
TACTGTTCACCATTCACAGCAGCTGCCAGGACCACCATGGAGGCTCTCCAGAAATGCCCATGA  
|||||  
X 10 20  
1520 1530 1540 1550 1560 1570 1580  
CTACAGCTGGCTCGGGGMAACCCCATGTCAATCAGGCGCTGAATTTGATCAGGTATGCAACAAGTTG  
|||||  
CTACAGCTGGCTCGGGGMAACCCCATGTCAATCAGGCGCTGAATTTGATCAGGTATGCAACAAGTTG  
30 40 50 60 70 80 90  
1590 1600 1610 1620 1630 1640 1650  
GGCTGGCTGAGCAATATTTGGATACATTTGCCCAACACAGATGTCCATTCATCATCACCATTCTCT  
|||||  
GGCTGGCTGAGCAATATTTGGATACATTTGCCCAACACAGATGTCCATTCATCATCACCATTCTCT  
100 110 120 130 140 150 160  
1660 1670 1680 1690 1700 1710 1720 1730  
GGCGCAGCTAACTTTGTAAGGCTGGCCCATAGATTCGAACCAATCAAGGGCTTCACAGTAATCT  
|||||  
GGCGCAGCTAACTTTGTAAGGCTGGCCCATAGATTCGAACCAATCAAGGG  
170 180 190 200 210 220 X

1740 1750 1760  
CCAAATTATCAGGCTCTGATCCAGTCCCTTCACAG  
7. US-09-301-380-1' (1-4134)  
aai50358 TOIG of: aai50358 check: 4462 from: 1 to: 611  
TOIG of: aai50358 check: 4462 from: 1 to: 611  
LOCUS AAI50358 611 bp mRNA linear EST 14-MAY-1997  
DEFINITION z107a03.r1 Soares pregnant uterus\_NbHPU Homo sapiens cDNA clone  
IMAGE:491596 5', mRNA sequence.  
ACCESSION AAI50358  
VERSION AAI50358.1 GI:1721870  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 611)  
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,  
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,  
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,  
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
and Maria,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
889549  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 1102 Std Error: 0.00  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 328.

FEATURES  
Location/Qualifiers  
1..611  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="GDB:3806183"  
/db\_xref="taxon:9606"  
/clone="IMAGE:491596"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="Soares\_pregnant\_uterus\_NbHPU"  
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
clig(dtr) primer [5',  
AACTGGAGAAATCGCGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bonaldo."

ORIGIN  
AAI50358 Length: 611 July 26, 2004 14:10 Type: N Check: 4462 ..  
Initial Score = 55 Optimized Score = 289 Significance = -1.50  
Residue Identity = 50% Matches = 391 Mismatches = 212  
Gaps = 175 Conservative Substitutions = 0  
2860 2870 2880 2890 2900 X 2910 2920  
TTTGGGTTGCCATTAGCTCTGCAGATCAAGGTCCTCTCTGGGACAGACACAGATTGTAGGGGC  
|||||

1470 1480 1490 1500 1510 1520 1530  
CAGGTTCCGAGGACCACCATGGGAGGTCTTCTCCAGAACTCCCATGATACA-GCT-GGCTCGGGGGCA  
CGAAANTCTAGAAATC---TGCA  
X 10  
X



```
1540 1550 1560 1570 1580 1590 1600
AACCCATGTCATTCAGGCGCTGAACCTTGTATCAGGTATG-GAACAAAGTTGGCGTGCCTGAGACAAATA
ATTCAAAGT--TCCAGAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
20 30 40 50 60 70 80

1610 1620 1630 1640 1650 1660 1670
TTTGGATACATTTGCCCAACACAGATGTCATCATCATCACCATCTTCTGGCGCCAGCTAACTTGTGA
NNNG-NNNAT---CC-----CNGANTAGATTC-----TCTGTGT-----CTAAGATTCTTA
90 100 110 120 130

1680 1690 1700 1710 1720 1730 1740
CTGAAGCGCTGGGCCATTAGATCGAAACCATTCGAAGGCTTCCACGTAATCTCCAAATATCAGGCTCTGA
--GATCAGATGCTCCA--AGATTCTAGATTAAA-----TA-----AGATTCTAACGGTCTGT
140 150 160 170 180

1750 1760 1770 1780 1790 1800 1810 1820
TCCAGTCTCTCACHAGCTGTGGGTTTTATCTGGTCTGAGGCTTTCGTCAAACTACTGCTCAGACGCTC
T--CTGT--TT-CAAGGCACTCTAGATTCCAT-TGGTCC--AAGATTCCG--GATCCT-----AAGNATC
190 200 210 220 230

1830 1840 1850 1860 1870 1880
GCTGGG-CAAGCTC---TCCCAATGCTGTCTACCTGCCATCACCGGAGAGTAGTTCACGTAAGAGACA
NAAGGANAGAGACNACACAGNNGGNNAGANANCAAGGNAATA-ANGTCNAANGNTGGACACC
240 250 260 270 280 290 300

1890 1900 1910 1920 1930 1940 1950
GCTTCAG-CTGGGCTGTGTCGTGT-TCCAGAAACTTCAGTTTGGTGGCCACACCCCTGGCTTGTGCAT
NNTAGGTCCTTCTGCAITCTGCTCTAGGACCAT-GGTTAAG-AGTCCAGAATCCACATTCTAAA-
310 320 330 340 350 360 370

1960 1970 1980 1990 2000 2010 2020
TGCACTTCAATA-TTCGATG--ATGAATTTGTAAGGGGCTATGTGTATCGCTGGGGTCCATGACAG
---ATCTT-ATAGTTCTAGCACTGTA-GTTCTAA-GACTCAAATGTTCTAAGTTCTAAGATTC-TAA-AG
380 390 400 410 420 430 440

2030 2040 2050 2060 2070 2080 2090 2100
CTGAACACTTTTCAAGTTGATCTGTAGTTCTAAGTCAAGGAGGATTTGGGACATCGTAAACGGGAGC
GTCCACA---GGTCTAG---ACTATTAGGTCAATTTCAAGG-----TCTAAC--CCTATACTGTAGT
450 460 470 480 490

2110 2120 2130 2140 2150 2160 2170
TGGAGTTGGAGTAGGACACACGCTTAACACAGCGCTGGCGGAGACGCTGTCCAGAGTGGTGTGGCCAC
ATTCTTTGGGGT--GOCCTCTCCTTCTTAGC-TATC-ATTGCTTCTCTC-CTCCCCA-ACTG----TGG----
500 510 520 530 540 550

2180 2190 2200 2210 2220 2230 2240
ACACGTGATAGTCCCGCTGTCACTGACATGACGTACACCTAGATG-ATCCCTGTCAACAG--TGAA
-----GGGT--GTGCCCT-TCA-AGCCTGTGCAAT--GC-ATTA-GGGATGCTCTCTTCCCGCAGGATGGA
560 570 580 590 600 610

2250 2260 2270 2280 2290 2300 2310
CCTTTCATCATCGGCGAGTCCCTGTGTGCTCTTACGCCA-CAGGACAGTGGGATAAGGTGTGATCATGTT
CGATTTT--CACCTTTC--GGGCCATG-TGCCCGGAGCATCCCGACCTTCTGAGA-NAGATGTGATC-TGGC
620 630 640 650 660 670 680

2320 2330 2340 2350 2360 2370
TCACTTTGCAATTCGAAGGACACCATGCTCCCTCTTTGCAACA CTGCAATTCGGGCTGTAAAGTGA
TCAAGGGGAGTCCA
690
X
```

> O <
O| | O Intelligence
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file 1\_x\_listing.res made by spaula on Mon 26 Jul 104 14:22:49-PDT.

Query sequence being compared:US-09-301-380-1 (1-4134)
Number of sequences searched: 8
Number of scores above cutoff: 8

Results of the initial comparison of US-09-301-380-1 (1-4134) with:
File : aal50358.seq
File : aal51738.seq
File : aak18454.seq
File : aak44359.seq
File : aak45811.seq
File : aba71469.seq
File : abs18595.seq
File : abs44016.seq

Table with 10 columns: N, U, M, B, E, R, O, F, S, E, Q, U, E, N, C, S. It contains a sequence alignment visualization with scores and gaps.

PARAMETERS

Table with 2 columns: Parameter, Value. Rows include Similarity matrix, Mismatch penalty, Gap penalty, Gap size, Cutoff score, and Randomization group.

SEARCH STATISTICS

Table with 2 columns: Statistic, Value. Rows include Scores (Mean, Median, Standard Deviation), Times (CPU, Total Elapsed), and Number of residues/sequences/scores.

The scores below are sorted by initial score.

Significance is calculated based on initial score.
A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Table with 5 columns: Sequence Name, Description, Length, Score, Init. Opt. Frame. It lists sequence details for aal50358 and aal51738.

1. US-09-301-380-1 (1-4134)
aal50358 TOIG of: aal50358 check: 4462 from: 1 to: 611

TOIG of: aal50358 check: 4462 from: 1 to: 611

LOCUS AAL50358 611 bp mRNA linear EST 14-MAY-1997
DEFINITION z107a03.r1 Soares\_pregnant\_uterus\_NbHPU Homo sapiens cDNA clone
IMAGE:491596 5', mRNA sequence.
ACCESSION AAL50358
VERSION AAL50358.1 GI:1721870
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 611)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfig,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
PUBMED 8889549

COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1102 Std Error: 0.00
Seq primer: -28M13 rev2 from AmerSham
High quality sequence stop: 328.

FEATURES
Location/Qualifiers
1..611
/organism="Homo sapiens"
/mol\_type="mRNA"
/db\_xref="GDB:3806183"
/db\_xref="taxon:9606"
/clone="IMAGE:491596"
/sex="female"
/dev\_stage="adult"
/lab\_host="DH10B"
/clone\_lib="Soares pregnant uterus NbHPU"
/note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15'

AA150358 Length: 611 July 26, 2004 14:10 Type: N Check: 4462 ..

Initial Score = 77 Optimized Score = 280 Significance = 0.98  
 Residue Identity = 49% Matches = 363 Mismatches = 237  
 Gaps = 130 Conservative Substitutions = 0

ORIGIN

```

AGAGTTTCTCAAGGTTTGAATGGGAGCCTTTATTTTCCATGTCTCCAGAGGACACCGCGAAGACTAT
700      710      720      730      740      750      760
|||||
AAAGAAC-CAGGCTTAG---AT
X
10

770      780      790      800      810      820      830
ATCTGTATGCTAGATTATATCACTACTCAACCATACAGACAGCAACCTATTCTGTGAAGGTGATTCA
|||||
TTTTTTTAAAG-TACA-TTGAATGTT-----GATGTTCACACAGAAC-ACCT-T-----AAG-TGTATACT
20      30      40      50      60      70

840      850      860      870      880      890      900
GTGGATGAATTGAATGACACTAGCTAGTCTTAAT-TTG-AGTGACACTAGTATTTATGGTCTAAATCAAGTA
|||||
GTTG-TG-----TAATGTCTCTAGAAAGG--AATCCTCTCTTAAACTGGGTTT--TGCTGTTTTCAGGTT
80      90      100      110      120      130

910      920      930      940      950      960      970
GAGAGAGGCCAACCAACATTTTAACTCCAGAAGGCA-ATGCAAGTAAACAAGAGGAATTAAAGAGGAATGTG
|||||
TCTA-----CCTAAATCATTTTT-----GGTATCTCTGATATCTCTA-TAA-TACTA-GAATGCT-
140      150      160      170      180      190

980      990      1000      1010      1020      1030      1040
CTTCTACTGAGTGCACTGAGAGGACTCCCTACCCCAATTTATTTACTGGCAAGAAAGATGGAATGCTA
|||||
CT-GCA-----AAATATAGT-AAGAAGAATTG--GAGCCTAAT-----AGCT-GATTCCT-
200      210      220      230

1050      1060      1070      1080      1090      1100      1110
CCCAAAAACAGGACAGTTTATAGAACTTTGAGAAAACCTTGAGATCATTCATGTTTCAGAACGACACTCT
|||||
CCCA-ATTTA--TCTG-TTAT-----GTTTGTCACTA--TT-CACAT-TTTAGTCTTT-----TC-TAGCAT
240      250      260      270      280      290

1120      1130      1140      1150      1160      1170      1180      1190
GGAAATTCCAAATGTA-TAGCAAAAATGCAATTAGAGGCAATCCACATACCATTTCTGTAGAGTTAAAGC
|||||
AAAAATTGTATGTGTACTTTCATGCCA-GTA-TAGGAACCTCAATC-TTTTTTTTTTTTCCCTTTAAGA
300      310      320      330      340      350

1200      1210      1220      1230      1240      1250      1260
GGCTCATACTGATCAGACCCCTCAAAATCTTGTG-CTGTCCCGAGAGAGATGGGACCTTGATC-TGC
|||||
AGGTTTTCAGT-GATTATA---CCTCAGGTATTTCTGAGTGT-CCTATTGCTCTAATAGGAGAATATCTTCC
360      370      380      390      400      410      420

1270      1280      1290      1300      1310      1320      1330
AGAGCT---AATGGCAACCCCAACCCAGATAGTGGTTAAACAATGAGTCCCAATAGAAATGCCCCCT
|||||
CGAGCTCAGAAATTAAGTTTC--TCTTA-AAATATAGAGATCCCAATCTTATGTAATA-ACCTAG-GCAT
430      440      450      460      470      480      490

1330      1340      1350      1360      1370      1380      1390      1400
GATGACCCAGCAGAAAATAGATGCGGATACCATTTATTTTCAAATGTTCAAGAAAGATCAAGTGCAGTA
|||||
GA-GTCCCTPAG-GGAGAAGTTAATGACCA-NTGGTTAAAGTGCC---TTTTTTAGAA--ATGCTGTGCGTGA
500      510      520      530      540      550

```

```

1410      1420      1430      1440      1450      1460      1470
TATCAGTGCATGCTCTTAATGAATATGGATATTACTGCA--AAGCATTTTGAATGCTGGCTGAGC
|||||
TGTC-CGAAATTGACATAAAGGATA--GACCTTGGCANGAGGAA---ATAAGTTAAAG
560      570      580      590      600      610

```

CACCACGAATCCTCACACCTGCAACACACTCTACCAG

2. US-09-301-380-1 (1-4134)  
 aak18454 Human brain expressed single exon probe SEQ ID NO:

TOIG of: aak18454 check: 7397 from: 1 to: 230

AAK18454 standard; DNA; 230 BP.

XX AAK18454;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe SEQ ID NO: 18445.

XX Human; brain expressed exon; gene expression analysis; probe; microarray;  
 XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;  
 XX ss.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains.

XX Example 4; SEQ ID NO 18445; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is one of the probes of the invention

XX Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;

XX AAK18454 Length: 230 July 26, 2004 14:10 Type: N Check: 7397 ..

Initial Score = 71 Optimized Score = 111 Significance = 0.68  
 Residue Identity = 52% Matches = 139 Mismatches = 84  
 Gaps = 42 Conservative Substitutions = 0

420 430 440 450 460 470 480 490  
 GGACGCTCATTAATTAACATCATGAGCGAAGGAAAGCTGAGACCTATGAGGAGTCTAT-CAGTGTACAGC



XX OS Homo sapiens.  
 XX PN WO200186003-A2.  
 XX PD 15-NOV-2001.  
 XX PF 30-JAN-2001; 2001WO-US0000665.  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PF 2002-114183/15.  
 XX PT Spatially-addressable set of single exon nucleic acid probes, used to  
 XX PT measure gene expression in human lung samples.  
 XX PS Claim 4; SEQ ID NO 18586; 634pp; English.  
 XX CC The invention relates to a spatially-addressable set of single exon  
 CC CC nucleic acid probes for measuring gene expression in a sample derived  
 CC CC from human lung comprising single exon nucleic acid probes having one of  
 CC CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC CC complements or the 12387 open reading frames derived from the 12614  
 CC CC probes. Also included are a microarray comprising the novel set of probes  
 CC CC; the novel set of probes which hybridise at high stringency to a nucleic  
 CC CC acid expressed in the human lung; measuring gene expression in a sample  
 CC CC derived from human lung, comprising (a) contacting the array with a  
 CC CC collection of detectably labeled nucleic acids derived from human lung  
 CC CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC CC algorithmically predicting at least one exon from genomic sequences of  
 CC CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,  
 CC CC having a fragment identical to the predicted exon, the probe is included  
 CC CC in the above mentioned microarray; assigning exons to a single gene,  
 CC CC comprising (a) identifying exons from genomic sequence by the method  
 CC CC above and (b) measuring the expression of each of the exons in several  
 CC CC tissues and/or cell types using hybridisation to a single exon  
 CC CC microarrays having a probe with the exon, where a common pattern of  
 CC CC expression of the exons in the tissues and/or cell types indicates that  
 CC CC the exons should be assigned to a single gene; a peptide comprising one  
 CC CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC CC analysis, and for identifying exons in a gene, particularly using human  
 CC CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
 CC CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC CC present sequence is a single exon probe open reading frame of the  
 CC CC invention. Note: The sequence data for this patent did not form part of  
 CC CC the printed specification, but was obtained in electronic format directly  
 CC CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX SQ Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;  
 ABS18595 Length: 230 July 26, 2004 14:11 Type: N Check: 7397 ..  
 Initial Score = 71 Optimized Score = 111 Significance = 0.68  
 Residue Identity = 52% Matches = 139 Mismatches = 84

Gaps = 42 Conservative Substitutions = 0  
 420 430 440 450 460 470 480  
 GGAACGCTCATAATTAACTATCATGAGCGAGGAAAGCTGAGACCTATGAGGAGTCTAT-CAGTGTACAGC  
 |||||  
 AGGAATTCATGCGCTTGACAGG  
 X 10 20  
 490 500 510 520 530 540 550  
 AAGGAACG-AACGCGGAGCTGCAGTTTCTAATAACATGTTGTCGGCCATCCAGATCAC--CATTTGGAC  
 |||||  
 A--GAAGTGCTCTGAGCTTGTGTTTC-CTTACGCGGCTCT-TTCTCTTTTACCACTGTATTGT---C  
 30 40 50 60 70 80  
 560 570 580 590 600 610 620  
 CAAGAAAACCTTGAACCAATCACACTTCAAGTGGTCACTTCTAGTCTTCTCTGAGACCCCAATTGG  
 |||||  
 C-AATAAAG---GAGCC-ATC-CTCATTTGAACCTGGCCA---TTAACCCCTT--CTCCATA--GTCAACTAG  
 90 100 110 120 130 140  
 630 640 650 660 670 680 690  
 ATTACCAACCACTTATATATTTGGATGATTAATTCCTTTC--AAAGACTTCCAAAGTGAGAGATTCT  
 |||||  
 GCTGCTGCA-CATCTCTCTTTT---TTCA-CAGTCTCTGCTGAAGGAGTTCGACTTCTCT-----TTTT  
 150 160 170 180 190 200  
 700 710 720 730 740 750 760 770  
 C-AAGGTTTGAATGGGAGCTTATTTTTCCTCAATGCTCTCCAGAGACACCCCGAGACTATATCTGTTA  
 |||||  
 CAAGGCTTG--TGGTCTTCTGATCA  
 210 220 230  
 TGCTA  
 5. US-09-301-380-1 (1-4134)  
 abs44016 Human liver single exon probe, SEQ ID No 19006.  
 TOIG of: abs44016 check: 7397 from: 1 to: 230  
 ID ABS44016 standard; DNA; 230 BP.  
 XX AC ABS44016;  
 XX AC ABS44016;  
 XX DT 25-FEB-2003 (first entry)  
 XX DE Human liver single exon probe, SEQ ID No 19006.  
 XX KW Human; single exon nucleic acid probe; liver; cirrhosis;  
 KW hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
 XX OS coronary heart disease; ss.  
 XX OS Homo sapiens.  
 XX PN WO200157273-A2.  
 XX PD 09-AUG-2001.  
 XX PF 30-JAN-2001; 2001WO-US0000664.  
 XX PR 04-FEB-2000; 2000US-0180312P.  
 XX PR 26-MAY-2000; 2000US-0207456P.  
 XX PR 30-JUN-2000; 2000US-00608408.  
 XX PR 03-AUG-2000; 2000US-00632366.  
 XX PR 21-SEP-2000; 2000US-0234687P.  
 XX PR 27-SEP-2000; 2000US-0236359P.  
 XX PR 04-OCT-2000; 2000GB-00024263.  
 XX PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX PF 2002-114183/15.

```
DR WPI; 2001-48898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PS gene expression in human adult liver.
XX
XX Claim 4; SEQ ID NO 19006; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pt_sequences
XX
SQ Sequence 230 BP; 45 A; 61 C; 44 G; 80 T; 0 U; 0 Other;
ABSA44016 Length: 230 July 26, 2004 14:11 Type: N Check: 7397 ..
Initial Score = 71 Optimized Score = 111 Significance = 0.68
Residue Identity = 52% Matches = 139 Mismatches = 84
Gaps = 42 Conservative Substitutions = 0
420 430 440 450 460 470 480
GGAAGCTCATATTAACATCATGAGCGAAGGAGCTGAGACCTATGAGGAGTCTAT-CAGTGTACAGC
|||||
X 10
490 500 510 520 530 540 550
AAGGAACG-AACGGGAGCTCAGTTCTTAATAACATGTTGTCGCCCATCCAGATCAC--CATTGTGGAC
|||||
A--GAAGTGCTCTGAGCTTCGTTTC-CTTCAGCCGGCTCT-TTCCTTTTACCACGTATTGT---C
30 40 50 60 70 80
560 570 580 590 600 610 620
CAAGAAACCTTGAACCAACACACTTCAAGTGGTCACTTTAGTACTTCTCGTGCAGACCCCAATTGG
|||||
C-AATAAG-----GAGCC-ATC-CTCATGAACTGGCCA---TTAACCCCTT--CTCCATA--GTCAACTAG
90 100 110 120 130 140
630 640 650 660 670 680 690
ATTACCAACCTATAATATTTTGGATGGATAATTCCTTTC--AAAGACTTCCACAAAGTGAGAGATTCTT
GCTGTGCTCA-CTAATCTCTTTT---TTCA-CAGTCCGTCTGGAAGGAGTTCGACTTCT-----TTTTT
150 160 170 180 190 200
700 710 720 X 730 740 750 760 770
C-AAGGTTTGAATGGGACCTTTATTTTTCATGTCCTCCAGAGGACCCCGGAAGACTATATCTGTGA
|||||
CAAGAGCTTG--TGGCTTCTGCAATCA
210 220 230
TGCTA
6. US-09-301-380-1 (1-4134)
aak45811 Human bone marrow expressed single exon probe SEQ
TOIG of: aak45811 check: 7602 from: 1 to: 224
ID AAK45811 standard; DNA; 224 BP.
XX
AC AAK45811;
XX,
```

```
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 20368.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN W0200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000668.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488900/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 20368; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is one of
XX the probes of the invention
XX
XX Sequence 224 BP; 58 A; 59 C; 49 G; 58 T; 0 U; 0 Other;
AAK45811 Length: 224 July 26, 2004 14:11 Type: N Check: 7602 ..
Initial Score = 38 Optimized Score = 107 Significance = -0.93
Residue Identity = 54% Matches = 134 Mismatches = 77
Gaps = 37 Conservative Substitutions = 0
2460 2470 2480 2490 2500 2510 2520
TACAAAGTTAGCTGGCGCCAGAAAGATGGTGAATGAATGACATCTGTGTTGTCGCAATGTATCCA--
|||||
TCTTCTCCAGAAATGT-CCCATG
X 10 20
2530 2540 2550 2560 2570 2580 2590
AATATA-TTGTCTCAGGCACGCCAA-CCTTTGT--TCCA-TACCTGATCAAGTTCAGCCCTGAATGACAT
|||||
ACTACAGCTGGCTCGGG--GGCAAAACCCCAATGTCATTCAGGGCCTGAACCTTGTACAGGTATGGA---ACA-
30 40 50 60 70 80
2600 2610 2620 2630 2640 2650 2660
GGGGTTTGC--CCCGAGCCAGCTGTAGTCAATGG-GACATTTCTGGAGAGACCTCCCAATGTGCTCTCTGG
|||||
AAGGTTGCGTGCCTGAGACA-ATATATT--TGGATACATTTGCCACAA-----CCACAGATG--TCCATT
90 100 110 120 130 140
2670 2680 2690 2700 2710 2720 2730
GAACGTGCGTGTGAATGGTGAACAG-TACCTTAG--CCGAGGTGCACTGGAGCCAGTACCTCTGAAAG
|||||
CATCATCACCATCTTCTGGCG-CCAGCTAACTTTGTACTGAAG-GC-CTGG--CCCATTAGAT--TCCGAAC
150 160 170 180 190 200 210
```

2740 X 2750 2760 2770 2780 2790  
 CATCCGAGGACACCTACAGGCTATCGGATTTACTATTGGAGACCCAGAGTTCATCTAA  
 |||||  
 CATTCAGGG  
 220 X

7. US-09-301-380-1 (1-4134)  
 aba71469 Human foetal liver single exon nucleic acid probe  
 TOIG of: aba71469 check: 7602 from: 1 to: 224

ID ABA71469 standard; DNA; 224 BP.  
 XX AC ABA71469;  
 XX AC  
 DT 01-FEB-2002 (first entry)  
 XX  
 DE Human foetal liver single exon nucleic acid probe #19774.  
 XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
 XX Homo sapiens.  
 XX  
 PN WO200157277-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US0000669.  
 XX  
 PR 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 XX WPI; 2001-483447/52.  
 DR  
 PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human foetal liver.  
 XX  
 PS Claim 4; SEQ ID NO 19774; 639pp + Sequence Listing; English.  
 XX  
 CC The invention relates to a single exon nucleic acid probe for measuring  
 CC human gene expression in a sample derived from human foetal liver. The  
 CC single exon nucleic acid probes may be used for predicting, measuring and  
 CC displaying gene expression in samples derived from human foetal liver. The  
 CC present sequence is a single exon nucleic acid probe of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 224 BP; 58 A; 59 C; 49 G; 58 T; 0 U; 0 Other;

ABA71469 Length: 224 July 26, 2004 14:11 Type: N Check: 7602 ..

Initial Score = 38 Optimized Score = 107 Significance = -0.93  
 Residue Identity = 54% Matches = 134 Mismatches = 77  
 Gaps = 37 Conservative Substitutions = 0

2460 2470 2480 2490 2500 2510 2520  
 TACAAAGTAGCTGGGCCCAAGAGATGGTATGATGATGGACATCTGGTGTGGCAAAATGTATCCA--  
 |||||  
 TCTTCTCCAGAATGT-CCCATG  
 X 10 20

2530 2540 2550 2560 2570 2580 2590

AAATATA-TTGCTCAGGCACGCCAA-CCTTTGT--TCCA-TACCTGATCAAAAGTTCAGGCCCTGAATGACAT  
 |||||  
 ACTACAGTGGCTCGGG--GGCAACCCCATGTCATTTCAGGCGCTGAACTTGATCAGTATGGA---ACA-  
 30 40 50 60 70 80  
 2600 2610 2620 2630 2640 2650 2660  
 GGGGTTTC--CCCGAGCCAGCTGTACTCATGG-GACATTCTGGAGAGACCTCCCAATGTGCTCCTGG  
 |||||  
 AAGGTGCGCTCGCTCGGACA-ATATATT--TGGATACATTTCGCCACAA-----CCACAGATG--TCCATT  
 90 100 110 120 130 140  
 2670 2680 2690 2700 2710 2720 2730  
 GAACGTGCGTGTGAATGTGTGACAG-TACCTTAG--CCGAGGTGCACTGGACCCAGTACCTCTGAAAG  
 |||||  
 CATCATCACATCTTTCTGGCG-CCAGCTAACTTTGTACTGAAG-GC-CTGG--CCCATTAGAT-TCGAAAC  
 150 160 170 180 190 200 210  
 2740 X 2750 2760 2770 2780 2790  
 CATCCGAGGACACCTACAGGCTATCGGATTTACTATTGGAGACCCAGAGTTCATCTAA  
 |||||  
 CATCAAGGG  
 220 X

8. US-09-301-380-1 (1-4134)  
 aa151738 TOIG of: aa151738 check: 8787 from: 1 to: 697

TOIG of: aa151738 check: 8787 from: 1 to: 697

LOCUS AA151738 697 bp mRNA linear EST 14-MAY-1997  
 DEFINITION z128h03.r1 Soares pregnant\_uterus\_NbHPU Homo sapiens cDNA clone  
 IMAGE:503285 5' similar to contains element MER22 repetitive  
 element ;, mRNA sequence.  
 ACCESSION AA151738  
 VERSION AA151738.1 GI:1720293  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 697)  
 AUTHORS Hillier,L., Lennon,G., Becker,M., Donald,M.F., Chiapelli,B.,  
 Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,  
 Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,M.,  
 Mardis,E., Moore,S., Morris,M., Parsons,J., Prange,C., Rifkin,L.,  
 Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,  
 Trevisan,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.  
 and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags  
 JOURNAL Genome Res. 6 (9), 807-828 (1996)  
 MEDLINE 97044478  
 PUBMED 8889549  
 COMMENT Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

This clone is available royalty-free through LML; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 813 Std Error: 0.00  
 Seq primer: -28M13 rev2 from Amersham  
 High quality sequence stop: 513.  
 Location/Qualifiers  
 1..697  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="GDB:3808273"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:503285"  
 /sex="female"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"

FEATURES  
 source

```
/clone lib="Soares_pregnant_uterus_NbHPU"  
/note="Organ: uterus; Vector: pTV73-Pac; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer [5',  
AACTGGAAGAATTGCGCGCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pTV73 vector. Library  
went through one round of normalization. Library  
constructed by M. Fatima Bernaldo."
```

## ORIGIN

AA151738 Length: 697 July 26, 2004 14:11 Type: N Check: 8787 ..

```
Initial Score = 25 Optimized Score = 294 Significance = -1.57  
Residue Identity = 46% Matches = 358 Mismatches = 317  
Gaps = 100 Conservative Substitutions = 0
```

```
10 20 30 40 50 60 70  
CTTCAAAGTCCCGCATGAAATTAATTAACGTTCACACACAGCTTCAGAAAATCTTTTGTGAAGAAG  
|||||  
CGAAANTCT-----AG  
X  
10  
80 90 100 110 120 130 140  
AAAAGGAAATTCAGTGTGTGAGTCTCAGCAG--GAGTTAAGCTAATGCAGCTTAAATAATGCCGAAAAGA  
|||||  
AATCTGCAATTCA-----AAAGTTCACAGAGNNNNNNNNNNAGGGGGNNNNNNNNNGG---GGGG  
20 30 40 50 60 70  
150 160 170 180 190 200 210  
AGCGTTATCTGGGGCAGAGTGGCGCTGA-TTCTCTTC-CTGTGCCAGATGATTAGTGCACTGGAAGTACC  
|||||  
NNNNNNNNNNNNNNNNNNATCCCGANNNTAGATTCTCTGTGC-TAAGATT-----CTAG-----ATC  
80 90 100 110 120 130  
220 230 240 250 260 270 280  
TCTTGATCCAAACTCTTTGAGAGTTGGTACAGCTCCACCACTCCAAACAGTCTCCAAAAGATTACAT  
|||||  
AGATGCTCCAAGA-TTCTAGATGATTAAATA-AGATTCTAACGGT--CTGTCTGT-TTCAAGGCAT-CTA  
140 150 160 170 180 190 200  
290 300 310 320 330 340 350  
TATTGACCTTCGGAGATATT-GTATCCAGTGTGAGCCAAAGGNAACCGCCCTTTCCTG--  
|||||  
GATT--CCATTGGTCCAGATTCGGATCC--TAAGNATCNAGGGANAAGCNCACNCGNNGGGNN  
210 220 230 240 250 260 270  
360 370 380 390 400 410 420  
---GACCGTAATGGGACTCATTTTGACATCGATAAGACC-CTCTGGTC-ACCATGAGCCTGGCACAGA  
|||||  
AGAGANNACAAAGGNAATAANGT--CNAANGNTGGACACCNNTAGGTCCTTGTGCAATCTGCCCTC-TCT  
280 290 300 310 320 330  
430 440 450 460 470 480  
ACGCTCATATTAACA-TC-ATGAGCGAAGGAAGCTGAGACCTATGAGGAGTCTA-TGRTGTACAGCA  
|||||  
AGGACCATGTTAAGAGTCCAAAGATCCA--CATTTCTAAATCT-TATAG--TTCTAGGCACGTGTATTTCT  
340 350 360 370 380 390 400  
490 500 510 520 530 540 550 560  
AGGACGAACGGGAGTGGAGTTCTTAATAACATTTGTGCCCCCATCCAGATCACCATTGGACCAA  
|||||  
AAGACTCAA--ATGTTCT-AAGTTCTTAAGATTCTAAAGTCCACAGTCTAG---ACTATT-AGGTGCAAT  
410 420 430 440 450 460 470  
570 580 590 600 610 620 630  
GAAAACCTGAACCAATCACATTCAAAGTGGTCTTTTAGTACTTCCCTGCAGACCCCAATTGGATTA  
|||||  
TTCAAGGTT---CTAA---CCCTACTAGTAGT-ATTCTTTT-GGGGTGCCCTCTCTCTTAGCTATCATG  
480 490 500 510 520 530  
640 650 660 670 680 690 700
```

```
CCACCACCTATAATATTTTGGATGGATTAATTCCTTTCAAAGACT-TCCACAAAGTGAGAGATTTCCTCAAGG  
|||||  
CTTCCTCTCCCAACTGTGG--GGTGTGCCCTTTC-AAGCCTGTGCAATGCATTAGGGA-TGCCTCTTTT  
540 550 560 570 580 590 600  
710 720 730 740 750 760 770  
TTTGAATG-GGGACCTTTT--ATTTTTC---CAATGCTCTCCAGAGGACACCCGGGAGACTATATCTGTTA  
|||||  
CCCGAGGATGGACGATTCACCTTTCCGGGCCATGTG-CCGGGA-GGCATCCCC---AC-CTTCTG---  
610 620 630 640 650 660  
780 790 800 X 810 820 830 840  
TGCTAGTTTATCATCTACTCAAAACCATACAGCAAGCAACCTATTTCTGTGAGGTGATTTCACTGGAAGA  
|||||  
AGANAGATGTGATCTGGCTCAAGGGGNAGTCCA  
670 680 690 X  
850  
ATTGATGACA
```